PATH-828 Bioinformatics for Cancer Research – Course outline 2018

Overarching objectives of the course include:

- Ascertaining associations/correlations between measures of gene expression and clinical or pathological parameters;
- Identify biomarkers using unsupervised (clustering) and supervised (classifiers, models) machine learning techniques as well as more widely used differential expression analysis;
- Survival analysis in relation to candidate biomarkers;
- Basic statistical analyses and power calculations to estimate required case numbers;
- Gene ontology / signaling pathways analysis;
- Appropriate pre-processing and data analysis techniques for various genetic data types such as microarray, tissue microarrays, methylation, NanoString, RNAseq, miRNAseq, proteomics and qRT-PCR

The course will introduce students to different types of data analysis. It will help them not only to perform analysis of their own data, but also give basic understanding of the statistical and data-mining/pattern recognition methods described in the literature.

It is **mandatory** for students to bring a laptop to each class.

Proposed topics covered by the course:

- Study design
    - Best practices for data collection and experiment designs.
    - How to avoid pitfalls of generating non-usable data
    - Showing how technical errors and poor data collection designs can affect the data analysis, using examples of sequencing, qRT-PCR and Nanostring data
    - Discussion and interpretation of plotted data

- Basic statistics for clinical and genetic research
    - meaning of the p-value, alpha, beta, hypothesis testing and effect size
    - statistical power – power calculations
    - univariate analysis, false discovery rate and fold change
    - correlation analysis and linear regression
    - survival analysis – understanding of the Kaplan-Meier plots and hazard ratio
    - resources to identify appropriate tests for a given data set (parametric vs non-parametric, continuous vs categorical variables, etc)
    - basics of multivariate statistical analysis

- Basic data-mining approaches and alternative methods to statistics for data analysis
    - Importance of data visualization with examples of good visualization techniques
    - Unsupervised learning – different types of clustering
    - Supervised learning
        - Feature selection
        - Basics on classifiers and prediction models
        - Validation (10-fold, leave 1 out)

- Gene ontology / signaling pathways analysis
    - Introduction
    - Exploring available tools for pathway analysis

- Analysis specifics and appropriate pre-processing steps for various genetic data types
    - Center for Advance Computing (CAC) resources
    - Publically available dataset resources (TCGA, ICGC, GEO)
    - microarray, tissue microarrays, methylation, NanoString, RNAseq, miRNAseq, proteomics and qRT-PCR, Next Generation Sequencing data analysis
        - Introduction to each
        - Exploring available tools/techniques

3 hour weekly lecture design

- First half or full lecture: topic-focused introductory lecture by course coordinator or an invited speaker
- Second half of the lecture: paper presentations (number of presentations will depend on the number of students enrolled)
    - Selected papers will illustrate valid or flawed statistical/data analysis methods
    - Data analysis of papers will reinforce topics covered in introductory lectures
    - All students will write a short critique of each weekly paper
    - One or two students will give a presentation of the paper followed by a group discussion

Homework

- Reading material to help students understand topic for the next week
- Small data analysis exercises to practice material covered so far
- Course Project
    - Project proposals are due mid-course.
    - The main project requirement is to apply at least two analysis methods to a genetic dataset. Students can use their own data or any publically available datasets.

Course Evaluation:

Presentations
Paper critiques
Class participation
Homework (online quiz)
Course project